DOCUMENT RESUME

ED 228 306 TM 830 228

AUTHOR Harnisch, Delwyn L.; Torres, Rosalie T.
TITLE Techniques for Detecting Student Errors: An Investigation with a Reading Test.
PUB DATE Apr 83
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Assessment; Elementary Secondary Education; Error Patterns; *Mathematics Skills; National Surveys; Reading Comprehension; *Reading Tests; *Responses; *Student Characteristics; Student Evaluation; Study Skills; Test Wiseness
IDENTIFIERS *Modified Caution Index; National Assessment of Educational Progress; *Response Patterns

ABSTRACT

An index identifying factors leading to unusual response patterns by students, the modified caution index, was computed for the reading comprehension and study skills items of Booklet 5 from the 1979 National Assessment of Educational Progress (NAEP) reading achievement survey given to the 9-, 13- and 17-year-old samples. The analyses focused on reporting student and school variables associated with unusual response patterns (high modified caution indices). The findings were compared to the results of a similar study which examined 13-year-olds' performance on the 1978 NAEP mathematics survey. Taken together, these studies provide a description of the relationship between the characteristics of students and their response patterns to a selection of items used to survey mathematics and reading performance levels in the United States. This description helps to illustrate how response pattern analyses conducted within local school settings can be used to identify individual students inadequately prepared for test material, as well as groups of students (classrooms) for whom content coverage was not consistent with specified instructional objectives. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED228306

Techniques for Detecting Student Errors: An Investigation with a Reading Test

Delwyn L. Harnisch and Rosalie T. Torres

University of Illinois, Urbana-Champaign

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D.L Harnisch

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830228

Paper presented at the annual meeting of the American Educational Research Association, Montreal, April 1983.

Abstract

This research is concerned with the identification of factors that lead to unusual response patterns to tests. Recently, several researchers have indicated the potential diagnostic value of appropriateness indices which could be evaluated at the student and school level along with student background characteristics. Specifically, indices based on the pattern of right and wrong answers were used to determine the unusualness of an individual's response pattern.

The data reported in this study are from the 1979-1980 NAEP reading survey administered to 9-, 13-, and 17-year olds. The sample consists of 2,627 9-year-old, 2,712 13-year-old, and 1,749 17-year-old students responding to reading comprehension and study skills. The analyses focused on reporting student and school variables associated with unusual response patterns. The findings consistent across age groups were those regarding (1) the extent of the unusualness of response patterns and (2) sex differences and the incidence of unusual response patterns. The findings inconsistent across age groups were those regarding the relationships between (3) reading performance level and the incidence of unusual response patterns and (4) ethnic differences and the incidence of unusual response patterns.

These results were compared to the findings of a similar study which examined 13-year olds' performance on the 1978 NAEP mathemtics survey. Taken together, these studies provide a description of the relationship between the characteristics of students and their response patterns to a selection of items used to survey mathematics and reading performance levels in the U.S. This description helps to illustrate how response

pattern analyses conducted within local school settings can be used to identify individual students inadequately prepared for test material, as well as groups of students (classrooms) for whom content coverage was not consistent with specified instructional objectives.

Techniques for Detecting Student Errors: An Investigation with a Reading Test

Introduction

Each year students in this country are administered millions of achievement tests. These tests are often used to provide a general assessment for individuals or large aggregates of students. Their results, whether for individual students, or for groups of students are usually reported in terms of a few summary scores based on the number of items answered correctly. While this practice is based upon the legitimate concern with the overall level of performance in a content area, it is clear that the same number-right score can be obtained in many different ways. In this sense focusing on summary scores can disquise important information contained in students' patterns of responses to individual test items.

Attention to the unusualness of response patterns can yield important information about student preparation. Students' responses are unusual to the extent that low scoring students (on total test score) answer some very difficult items correctly, and high scoring students answer some very easy items incorrectly. Among the wide variety of factors that can lead to unusual response patterns are guessing, carelessness, high anxiety, an unusual instructional history or other experiential background, a localized misunderstanding that influences responses to a subset of items, or copying a neighbor's answers to certain questions.

An important factor in using response patterns to interpret test results is illustrated by considering a short test of just five items.

On such a test a score of three right answers can be obtained by ten different combinations of correct and incorrect responses. The number of possible combinations expands rapidly with increases in test length. Obviously it is not feasible to provide different interpretations for each unique pattern of responses.

However, recent developments of powerful and relatively simple techniques for identifying atypical response patterns has made this information much more readily available. These techniques have been used for a number of purposes with students' responses to items on a mathematics achievement test (Harnisch & Linn, 1982). The purpose of this study is to extend this use to reading achievement test data.

Use of Response Pattern Analyses

The information made available by response pattern analyses has several possible uses. The most obvious is simply that of identifying individual students or groups of students for whom special caution is needed in interpreting their total correct scores. Secondly, the content characteristics of items that contribute to a high caution index for a students can be examined, and may help to identify particular strengths or gaps in a student's knowledge. In addition, examination of the content characteristics of these items may help to identify a mismatch between instructional practices and test content. Thirdly, then, between school differences in response patterns can be identified and associated with differences in instructional practices or other school characteristics.

Harnisch and Linn (1982) examined student test data from the 1978 National Assessment of Educational Progress (NAEP) mathematics survey given to 13-year olds and found (1) unusual response patterns associated

with low total scores; (2) blacks and males to have a higher incidence of unusual response patterns than whites and females, respectively; and (3) sizeable between school differences in the unusualness of response patterns. A further question is whether or not results such as these are particular to mathematics achievement or to a particular age group. The present study addresses this question by extending response pattern analyses to NAEP reading achievement data for 9-, 13-, and 17-year olds. Prior to describing this study we will present a brief description of the recently developed techniques used to analyze response patterns.

Definition of Indices

Several indices based upon students' patterns of responses to individual items have been defined. These include the "caution" index proposed by Sato (1975), the modified caution index by Harnisch and Linn (1981), the U' index by van der Flier (1977), the personal biserial by Donlon and Fischer (1968), the norm conformity index by Tatsuoka and Tatsuoka (1980), and the agreement and disagreement indices by Kane and Brennan (1980). These indices have been termed "group dependent indices" because their calculation depends upon the total group of students for whom response patterns are being analyzed.

For the purposes of understanding the nature of these indices it is convenient to start with a consideration of a matrix of zeroes and ones as shown in Table 1. A row of the matrix is associated with each examineee and a column with each item. Ones are recorded for correct responses and zeros for incorrect responses. Rows and colums of the data matrix are permuted so that the items (columns) are arranged, from left to right in ascending order of difficulty, and examinees (rows) are arranged, from top to bottom in descending order of total number of

correct answers. The resulting matrix has been called an "S-P" Table (student-problem table) by Sato (1975). (See Tatsuoka (1978) for a description in English.)

Insert Table 1 about here

If the items formed a perfect Guttman scale (Guttman, 1941), the S-P Table would consist of a section with all ones in the upper left-hand corner and all zeroes in the lower right-hand corner. A single step-like boundary line would separate the ones and the zeroes. In other words, anyone who responded correctly to a difficult item would also answer all easier items correctly. There would be no unusual response patterns (in the sense that is used for the indices described in this paper) because everyone who had a given total score would have the same pattern of responses. Of course, with responses to achievement test items, perfect Guttman scales are not expected. Consequently, the S-P Table will be characterized by a predominance of ones in the upper-left hand corner and zeroes in the lower right-hand corner, but there will be many exceptions to the pattern, i.e., ones in the region where mostly zeroes are found and vice versa.

A small hypothetical example of an S-P Table with 15 examinees and 10 items is shown in Table 1. The solid and dashed lines in Table 1 are known as the S-curve and P-curve respectively. The S-curve (solid line) is obtained by drawing a vertical line for each row that has $n_{i.}$ items (columns) to the left of it where $n_{i.}$ is the total number of correct responses for the ith examinee. The P-curve (dashed line) is obtained by

drawing a horizontal line in each column such that there are n_j examinees above the horizontal line who answer item j correctly.

For an ideal, or Guttman-scalable S-P Table, the S- and P- curves would coincide. The degree of divergence indicates heterogeneity while the degree of convergence indicates homogeneity of response patterns.

Sato's (1975) use of the name, caution index comes from the notion that a large value is associated with examinees that have unusual response patterns. It denotes that some caution may be needed in interpreting a total correct score for an examinee. The key point is that the caution index provided information about an examinee that is not contained in the total score. A large value of the caution index raises doubts about the validity of the usual interpretation of the total score for an individual.

Sato's caution index can be computed for items as well. Similarly, large values of the caution index for items reflects the fact that either a very easy item was missed by high scoring students or a very difficult item was answered correctly by low scoring students. A high caution index for an item indicates that it is an item which is contributing to high caution indices for students. The contents of items with high caution indices should be examined in order both to help identify possible mismatches between instructional practices and test content, and to judge the quality of the item's wording.

A modified form of Sato's caution index, C_i^*, was introduced by Harnisch and Linn (1981) to yield a lower bound of 0 and upper bound of 1. (Modified caution indices for students and items are computed and listed for the 15 students and 10 items in Table 1.) By being bound by 0 and 1, the modified caution index eliminates the extreme scores that are

sometimes obtained on the caution index, especially in cases where a very high scoring examinee misses a single very easy item. For this reason the analyses of the present study focus on the use of the modified caution index as an indicator of the unusualness of response patterns.

Calculation of the remaining group dependent indices is based upon the S-P Table as well. Harnisch and Linn (1981) describe the computational formulas for all of the indices, and found these indices to be highly intercorrelated.

Methodology

Data Source

The data used in this investigation came the 1979-1980 NAEP reading survey of 9-, 13-, and 17-year olds. During the 1979-80 school year, the NAEP surveyed over 100,000 randomly selected students using 40 different booklets containing items addressing their attitudes and achievement in reading and literature. Approximately 2,700 students were randomly selected for each booklet.

The data for this analysis came from Booklet 5 given to 9-, 13-, and 17-year olds. (Nine exercises contained identical content for the 9- and 13-year olds, while twelve exercises were identical for the 13- and 17-year olds. No items of identical content were administered to the 9- and 17-year olds in Booklet 5.) Booklet 5 was selected primarily because all of the items in each Booklet 5 across the three age groups have been publicly released; and secondly, on the basis of the quality of the attitudinal items. Public use data files containing the responses to Booklet 5 at each age level were purchased from the Education Commission of the States (ECS) (Note 1).

NAEP Sampling. A deeply stratified, multi-stage sampling design with oversampling of low-income and rural areas was used by NAEP (National Center for Educational Statistics, 1978; Moore, Chromy, & Rogers, 1974). In the first stage, the United States is divided into geographical units, including Standard Metropolitan Statistical Areas and counties or groups of contiguous counties. These are stratified according to region and size of community. In the second stage, schools are randomly sampled within selected first-stage geographical units. Finally, from 10 to 35 age-eligible students are randomly selected within schools for testing.

NAEP Data for Secondary Analysis

The NAEP test materials are organized into booklets suitable for administration in a single sitting. Each booklet or "test package" contains a collection of items taken by a selected student. In the 1979-1980 reading assessment, 11, 15, and 14 separate booklets were each administered to large national samples of 9-, 13-, and 17-year olds, respectively.

The public use data tape provided by the NAEP contains information on 2,627 9-year olds; 2,712 13-year olds; and 2,749 17-year olds responding to questions on Booklet 5. The characteristics of each age group are similar. The sample is composed of an almost equal number of boys and girls. Most students (between 90.5% at age 13 and 94.6% at age 9) were attending a public school as compared to a private school. The majority of the students (between 77.6% at age 13 and 80.4% at age 17) are white, while blacks and other racial groups are represented. Most parents of students (between 63.8% at age 9 and 82.1% at age 17) are either high school graduates or had attended college.

Present Analysis

Weighting. The NAEP sampling procedures require that individual weighting procedures be used to reflect the representative target populations. In order to obtain unbiased estimates of population means we took the individual weight value supplied by NAEP on the tape file and created a weighted data file with observations being included in the sample. All correlational analyses were based on the weighted data file for the sampling design, and therefore the results apply to a representative population of students in the United States (Moore et al., 1974).

Items selected for analysis. The 1979-1980 NAEP reading survey addressed the following objectives: valuing reading and literature, comprehending both literary and expository reading works, responding to written works, and applying study skills. Each booklet contains items which assess each objective. The items selected for the present analysis assessed two of the objectives, reading comprehension and study skills. The items related to these objectives were selected because it was felt that they most closely assessed mastery of reading skills. There were 23, 25, and 21 reading comprehension and study skills items in Booklet 5 for the 9-, 13-, and 17-year olds, respectively. For the purposes of the present analysis total test score refers to these selected items only.

Procedure. The data of the present analysis were examined with respect to (1) the difficulty level of the items; (2) the intercorrelations among the various group dependent indices; (3) the relationships between students' reading performance, sex, race, and the modified caution index; (4) the characteristics of students who have high modified caution indices; and (5) the characteristics of schools associated with high modified caution indices.

Results and Discussion

Item Difficulties and Reading Performance

To illustrate the varying distributions of item difficulties for each age group we prepared a "Box and Whisker Plot." A box and whisker plot provides a graphical picture of the location of the 25th, 50th, and 75th percentile score with the range of values plus the mean value. The box represents the middle 50% of the values while the whisker or tails extending from the box represent the lower and upper quartile values. Values of 0 on the box and whisker plot indicate the probability occurrence as 1 chance out of 20, while an asterisk indicates the probability of occurrence as 1 chance out of 200.

A box and whisker plot of the item difficulties for the reading comprehension and study skills items of Booklet 5 for the 9-, 13-, and 17-year olds is given in Figure 1. The range of item difficulties for the 13- and 17-year olds are approximately equal and somewhat higher than the range of item difficulties for the 9-year olds. Thus, as indicated by the plus signs within each box, the mean percentage of reading comprehension and study skills items answered correctly by the 13- and 17-year olds (74.4% and 76.0%) is higher than the mean percentage of items answered correctly by the 9-year olds (58.7%). The distributions of total test scores for the reading comprehension and study skill items for each age group are all negatively skewed (skewness = -.204, -.986, and -.764 for the 9-, 13-, and 17-year olds, respectively) as would be expected given the above range of item difficultie for the three age groups.

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

Response Pattern Analyses

Modified caution indices computed for each group increase slightly in both size and variability with age. The mean modfied caution indices are .22, .25, and .27 for the 9-, 13-, and 17-year olds respectively; while the standard deviations are .14, .17, and .20.

Intercorrelations among group dependent indices. Intercorrelations between Sato's caution index, the modified caution index, the six other group dependent indices, and the total test score were computed for each age group. The results are consistent with those found with the analysis of the 1977-1978 NAEP mathematics survey of 13-year olds (Harnisch & Linn, 1982). The correlation among Sato's caution index and the modified caution index is .97 or higher for all three age groups. The correlations of all the group dependent indices with total test scores vary from being moderately negative (-.48) to almost perfect (.99). These correlations are consistent with previous research that suggests that the sign and magnitude of the correlation between total scores and group dependent indices varies as a function of the total test score distribution (Harnisch & Linn, 1981).

Modified caution index and reading performance. The modified caution index is negatively related (-.23) with total test score for the 9-year olds, but unrelated (-.04 and .07) for the 13- and 17-year olds. This result for 9-year olds is similar to the small negative correlation (-.31) found between the modified caution index and 13-year olds total test scores on the 1977-1978 NAEP mathematics survey (Harnisch & Linn, 1982). The presence of high caution indices for individuals with low total scores could possibly result from a combination of factors

including (1) guessing correctly on one or more of the very difficult items; (2) irregular study habits for low achievers; (3) long-term absence from instruction; (4) a wide range of item difficulties on the test; and/or (5) teachers emphasizing study skill areas found on the NAEP survey which for the majority of students may be more difficult than for this particular group, or conversely, teachers not covering parts of the curriculum that are in the NAEP reading surveys. For example, 9-year olds were found to have poorer performance on items requiring the comprehension of poems than 13-year olds. Their poor performance may be due to the fact the 9-year olds are not familiar with poetry as a literary form.

Modified caution index and sex. Box and whisker plots of the modified caution index for males and females by their test item performance are given in Figures 2, 3, and 4 for the 9-, 13-, and 17-year olds, respectively. The total test score interval for each age group was split into quintiles, and lines were drawn between the subgroup means at each ability level on the box and whisker plots. For each ability level of each age group the mean and median of the modified caution index is the same or greater for males than for females as shown in Figures 2, 3, and 4. The difference between males and females in the central tendency of the modified caution index distribution is seen in the slope of the line connecting the box and whisker plots at each ability level. That is, a negative slope for the line connecting the average modified caution index values indicates a greater value for males than for females. Significant differences ($p < .01$) were found between the mean modified caution indices of males and females for each age group. Again, this

result is conistent with the previous study examining 13-year-old math performance (Harnisch & Linn, 1982).

- - - - - - - - - - - - - - - - - - - -

Insert Figures 2, 3, and 4 about here

- - - - - - - - - - - - - - - - - - - -

Caution index and race. Similar analyses were conducted for each age group by race. Significant differences were found between blacks and whites for the 9- and 17-year olds. While for the 9-year olds, blacks had significantly higher caution indices than whites ($p < .001$); for the 17-year olds, whites had significantly higher caution indices than blacks ($p < .001$). This result for 9-year olds is consistent with Harnisch and Linn's (1982) finding of higher caution indices for 13-year-old blacks than whites in math performance. A breakdown of these differences by quintiles revealed the greatest differences to be between 17-year old whites and blacks who scored in the second and third quintiles for total test score. These findings suggest that the test content emphasized may be more relevant (1) for 9-year old whites than 9-year old blacks and (2) for 17-year old blacks than 17-year old whites.

Characteristics of students with high modified caution indices. Selected characteristics of 9-year olds with high modified caution indices ($\geq .30$) were examined in order to illustrate analyses which relate unusual response patterns to student characteristics. These characteristics included both students' background characteristics and students' attitudes toward reading. Students' background characteristics described their region of the country, their community size, whether or not English is spoken most often in their homes, and whether or not they attended kindergarten. Students' attitudinal items described their

opinions toward the importance of reading, their reading ability, their enjoyment of reading, and reading in their spare time.

The percentages of students with high modified caution indices responding in each category of these items are given in Table 2. For example, considering Item 1 of Table 2, it can be seen that of the students who live in the Northeast, 27.3% have high modified caution indices. With respect to background characteristics, the greatest percentages of students with high modified caution indices (1) live in the Southeast (29.6); (2) live in big cities (28.0) and in smaller places (26.7); (4) come from homes where English is not the language spoken most often (30.9); and (4) did not go to kindergarten before attending first grade (34.0). With respect to attitudes toward reading, the greatest percentages of students with high modified caution indices (1) feel that reading is not important at all (46.2); (2) feel that they are very poor readers (45.3); (3) do not enjoy reading all (36.5); (4) never read for their own enjoyment during spare time (36.2); and (5) mostly read stories about real people, places, or things in their spare time (29.2)

- - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - -

Characteristics of schools associated with high modified caution indices. Similarly, the characteristics of schools associated with high modified caution indices for 9-year olds were examined in order to illustrate analyses which relate unusual response patterns to school characteristics. These analyses were conducted by computing mean caution indices for the 9-year olds of each school and then comparing the characteristics of schools with high modified caution indices ($\geq$ .30) to

those with low modified caution indices (< .30). School associated with high modified caution indices were found to have populations of significantly ($p < .0001$) (1) fewer Asian students and (2) more black students; and to be located in communities with significantly ($p < .0001$) (1) more farm workers, (2) fewer professional workers, and (3) more families on welfare.

Although these background, attitudinal, and school characteristics would typically be expected to describe students with poor reading performance and might be thought to reflect the small negative correlation (-.23) between the modified caution index and total test score for the 9-year-olds, a similar result was found when total test performance was controlled for in examining these differences.

Summary and Implications

An index describing the unusualness of students' response patterns, the modified caution index, was computed for the reading comprehension and study skills items of Booklet 5 from the 1979 NAEP reading achievement survey given to the 9-, 13-, and 17-year old samples. The analyses focused on reporting student and school variables associated with unusual response patterns (high modified caution indices). The firdings were compared to the results of a similar study which examined 13-year olds' performance on the 1978 NAEP mathematics survey.

The modified caution indices computed for each age group were found to be similar. A small negative correlation was found between the modified caution index and reading performance for the 9-year olds, while near zero correlations were found for the 13- and 17-year olds.

With respect to sex, males of all performance levels of each age group were found to have higher modified caution indices than females.

With respect to race, 9-year-old blacks were found to have higher modified caution indices than 9-year-old whites, while 17-year-old whites were found to have higher modified caution indices than 17-year-old blacks.

Thus, while some findings were consistent across age groups, others were not. The consistent findings were those regarding (1) the extent of the unusualness of response patterns and (2) sex differences and the incidence of unusual response patterns. The inconsistent findings were those regarding the relationships between (3) reading performance level and the incidence of unusual response patterns and (4) ethnic differences and the incidence of unusual response patterns.

Of these findings, however, several are similar to those found with 13-year olds' mathematics performance (Harnisch & Linn, 1982). First, a small negative correlation was found between 13-year olds' mathematics performance and the modified caution index. In this study, the same relationship was found between 9-year olds' reading performance and the modified caution index. Second, 13-year-old males were found to have higher modified caution indices than 13-year-old females on mathematics performance. This study found the same relationship between males and females of all age groups on reading performance. Third, 13-year-old blacks were found to have higher caution indices on mathematics performance that 13-year-old whites. In this study, the same relationship was found for 9-year-old blacks and whites on reading performance.

Taken together, these findings suggest that the relationship between unusual response patterns and sex, that is, males having more unusual response patterns than females, may be consistent across age levels and

content areas. The stability of this relationship is further supported by various studies which found (1) boys to show more curiosity and exploratory behavior than girls (Hutt, 1970; Maw & Maw, 1965; Shipman, 1971); (2) boys to show more originality in nonverbal of creativity Torrance, 1965; (Torrance & Aliotti, 1969); (3) preschool boys to prefer unstructured activities (Carpenter & Huston-Stein, 1980), and (4) boys to less frequently select "I don't know" responses to NAEP science achievement items (Linn, Benedictis, Delucci, Stage, & Harris, 1982). These findings suggest that boys may tend to engage in more risk-taking behavior than girls, and thus, might be expected to engage in behaviors (e.g., guessing) resulting in a higher incidence of unusual response patterns.

On the other hand, less consistency was found between unusual response patterns and both test performance and race across age levels and content areas. This finding suggests that the incidence of unusual response patterns may be quite situation specific. That is, as suggested by Harnisch and Linn's (1982) finding of sizeable between school differences in modified caution indices, significant variation may exist between instructional settings. Differences in instructional settings may in turn contribute to differences in factors which contribute to unusual response patterns. These factors include variation in emphasis in content coverage, or in the instructional history or experiential backgrounds of students across different age levels, content areas, and/or ethnic groups.

Finally, in order to illustrate analyses which relate unusual response patterns to student and school characteristics; selected

background, attitudinal, and school characteristics of 9-year olds with high modified caution indices were examined.

The following background characteristics of 9-year olds were found to be associated with the greatest percentages of students who had high modified caution indices: living in the Southeast, living in big cities or in smaller places, being from homes where English is not the predominant language, and not having attended kindergarten. In addition, the greatest percentages of 9-year olds with high modified caution indices were found to have poor attitudes toward reading (e.g., not feeling that reading is important, feeling that they are poor readers, etc.).

Schools whose 9-year-old students had high modifed caution indices were found to have fewer Asian and more black students; and to be located in communities with more farm workers, fewer professional workers, and more families on welfare. These students and school characteristics were found to describe students of differing reading performance levels, and, thus, do not merely reflect a relationship between poor reading performance and the incidence of unusual response patterns.

In summary, then, the research conducted on response patterns thus far has provided a description of the relationship between the characteristics of students and their response patterns to a selection of items used to survey mathematics and reading performance levels in the U.S. While on the one hand, this description has shown variability in the incidence of unusual response patterns that would be expected as a result of variability in content coverage; on the other hand, it illustrates how response pattern analyses conducted within a local school district can be productive on at least two levels. These analyses

conducted within a single setting where the content of instruction is based upon a specific set of objectives can be used to identify (1) individual students who are inadequately prepared for test material, and (2) groups of students (classrooms) for whom content coverage was not consistent with agreed-upon objectives expected to be addressed by teachers.

Further, conducting these analyses by subsets of items designed to assess subobjectives can provide even more detailed information about school and classroom differences in student performance. Such an analysis for the reading achievement data of the present study could possibly identify the subset of items which contribute most to the caution indices for schools or classrooms with high means on the caution indices. The analysis of subsets of items could be used to describe unique patterns of performance by item content. Furthermore, inconsistent patterns suggest differences in coverage that make the test less appropriate for some schools than others. These results can serve a variety of uses: (1) identifying school and classroom strengths and weaknesses with respect to content coverage, (2) improving inservice teacher education, and (3) improving the design and development of reading comprehension tests.

Note

1. Public Use Data Tapes can be purchased from:

   Department of User Services
   National Assessment of Educational Progress
   Education Commission of the States
   1860 Lincoln Street, Suite 700
   Denver, Colorado 80295

References

Carpenter, C. J., & Huston-Stein, A. Activity structure and sex-typed behavior in preschool children. Child Development, 1980, 51, 862-872.

Donlon, T. F., & Fischer, F. E. An index of an individual's agreement with group-determined item difficulties. Educational Psychological Measurement, 1968, 28, 105-113.

Guttman, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst, P. Wallin, & L. Guttman (Eds.), The prediction of personal adjustment. New York: Social Science Research Council, Committee on Social Adjustment, 1941.

Harnisch, D. L., & Linn, R. L. Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, 1981, 18, 133-146.

Harnisch, D. L., & Linn, R. L. Identification of aberrant response patterns in a test of mathematics. Paper presented at the annual meeting of the American Educational Research Association, New York, March 1982.

Hutt, C. Curiosity in young children. Science Journal, 1970, 6, 68-71.

Kane, M. T., & Brennan, R. L. Agreement coefficients as indices of dependability for domain-referenced tests. Applied Psychological Measurement, 1980, 4, 105-126.

Linn, M., Benedictis, T., Delucci, K., Stage, E., & Harris A. Sex Differences in Science: "I don't know". Paper presented at the 1982 annual meeting of the American Educational Research Association, New York, March 1982.

Maw, W. H., & Maw, E. W. Differences in preference for investigatory activities by school children who differ in curiosity level. Psychology in the Schools, 1965, 2, 263-266.

Moore, R. P., Chromy, J. R., & Rogers, W. T. The national assessment approach to sampling. Denver, CO: Education commission of the States, National Assessment of Educational Progress, 1974.

National Center for Education Statistics. Three national assessments of science: Changes in achievement, 1969-77. Science Report No. 08-5-00, Denver, CO: National Assessment of Educational Progress, 1978.

Sato, T. The construction and interpretation of S-P tables. Tokyo: Meiji Tosho, 1975.

Shipman, V. C. Disadvantaged children and their first school experiences. Educational Testing Service Head Start Longitudinal Study, 1971.

Tatsuoka, K., & Tatsuoka, M. M. Detection of aberrant response patterns and their effects on dimensionality. (Research Report 80-4) Urbana, IL: University of Illinois, Computer-based Education Research Laboratory, 1980.

Tatsuoka, M. M. Recent psychometric developments in Japan: Engineers grapple with educational measurement problems. Paper presented at the ONR Contractors Meeting on Individualized Measurement, Columbia, MO, 1978.

Torrance, E. P. Rewarding creative behavior. Englewood Cliffs, NJ: Prentice-Hall,1 1966.

Torrance, E. P., & Aliotti, N. C. Sex differences in levels of performance and test-retest reliability on the Torrance Tests of Creative Thinking Ability. Journal of Creative Behaviors, 1969, 3, 52-57.

van der Flier, H. Environmental Factors and deviant response patterns. In Y. H. Poortinga (Ed.), Basic problems in cross-cultural psychology. Amsterdam: Swets and Seitlinger, B. V., 1977.

Table 1

S-P Table for 15 Students and 10 Items
(Hypothetical Example)

Examinee i	Item j 2	8	1	3	4	6	10	9	5	7	Examinee Total $n_{i.}$	Modified Caution Index C_i^*
7	1	1	1	1	1	1	1	1	1	1	10	.00
5	1	1	1	1	1	1	1	1	1	0	9	.00
9	1	1	1	1	1	0	1	1	0	1	8	.29
4	1	1	1	1	1	1	1	0	0	0	7	.00
2	1	1	0	1	1	0	1	1	0	0	6	.25
10	1	1	1	0	1	0	1	0	1	0	6	.25
13	1	1	0	1	0	0	1	0	0	1	5	.35
1	1	1	1	0	1	0	0	0	1	0	5	.20
6	1	0	0	1	0	1	0	1	1	0	5	.50
14	1	1	1	1	0	1	0	0	0	0	5	.00
11	1	0	0	0	0	1	0	1	0	1	4	.60
15	0	1	1	1	0	1	0	0	0	0	4	.20
3	1	0	1	0	1	0	0	0	0	0	3	.17
8	0	0	1	0	0	1	0	0	0	0	2	.36
12	0	1	0	0	0	0	0	0	0	0	1	.12
Item Total $n_{.j}$	12	11	10	9	8	8	7	6	5	4		
Modified Caution Index C_j^*	.05	.17	.27	.07	.07	.37	.00	.15	.19	.29		

Table 2

Percentage of Students in each Category of Selected Background and Attitudinal Items with High Modified Caution Indices ($\geq$ .30)
Age 9, NAEP Booklet 5

Background Items	%
1. Region of Country	
Northeast	27.3
Southeast	29.6
Central	19.7
West	24.9
2. Community Size	
Big city	28.0
Fringes of big city	19.9
Medium city	23.1
Smaller places	26.7
3. Is English the language spoken most often in your home?	
Yes	23.2
No	30.9
4. Did you go to kindergarten before you attended first grade?	
Yes	23.5
No	34.0

Attitudinal Items	
1. How important is it to be able to read?	
Very important	24.4
Somewhat important	34.8
Not important at all	46.2
2. What kind of reader do you think you are for your age?	
A poor reader	45.3
A good reader	24.1
A very good reader	25.4
3. How much do you enjoy reading?	
Very much	24.3
Somewhat	27.4
Not at all	36.5
4. How often do you read for your own enjoyment during your spare time?	
Almost every day	23.5
Once or twice a week	25.6
Less than once a week	28.2
Never	36.2
5. Which statement best describes the type of reading you do in your spare time?	
Mostly read 'make believe' stories	23.7
Mostly read stories about real people, places or things	29.2
Read almost the same number of make believe and real stories	22.0
Never read in spare time	32.9

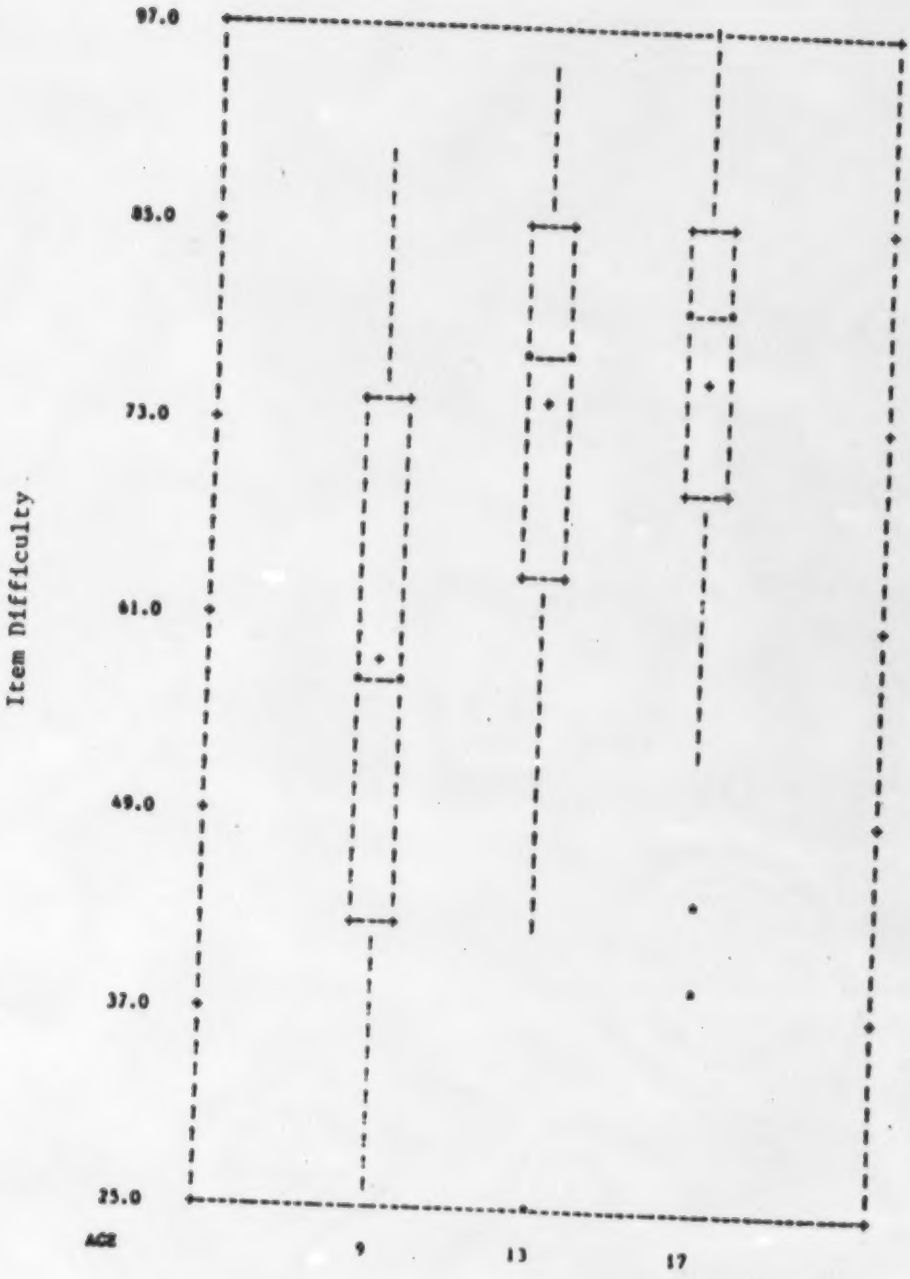

Figure 1. Box and whisker plot of item difficulties for each test booklet administered to 9-, 13-, and 17-year olds. (NAEP Booklet 5, Study skills and reading comprehension items, 1979)

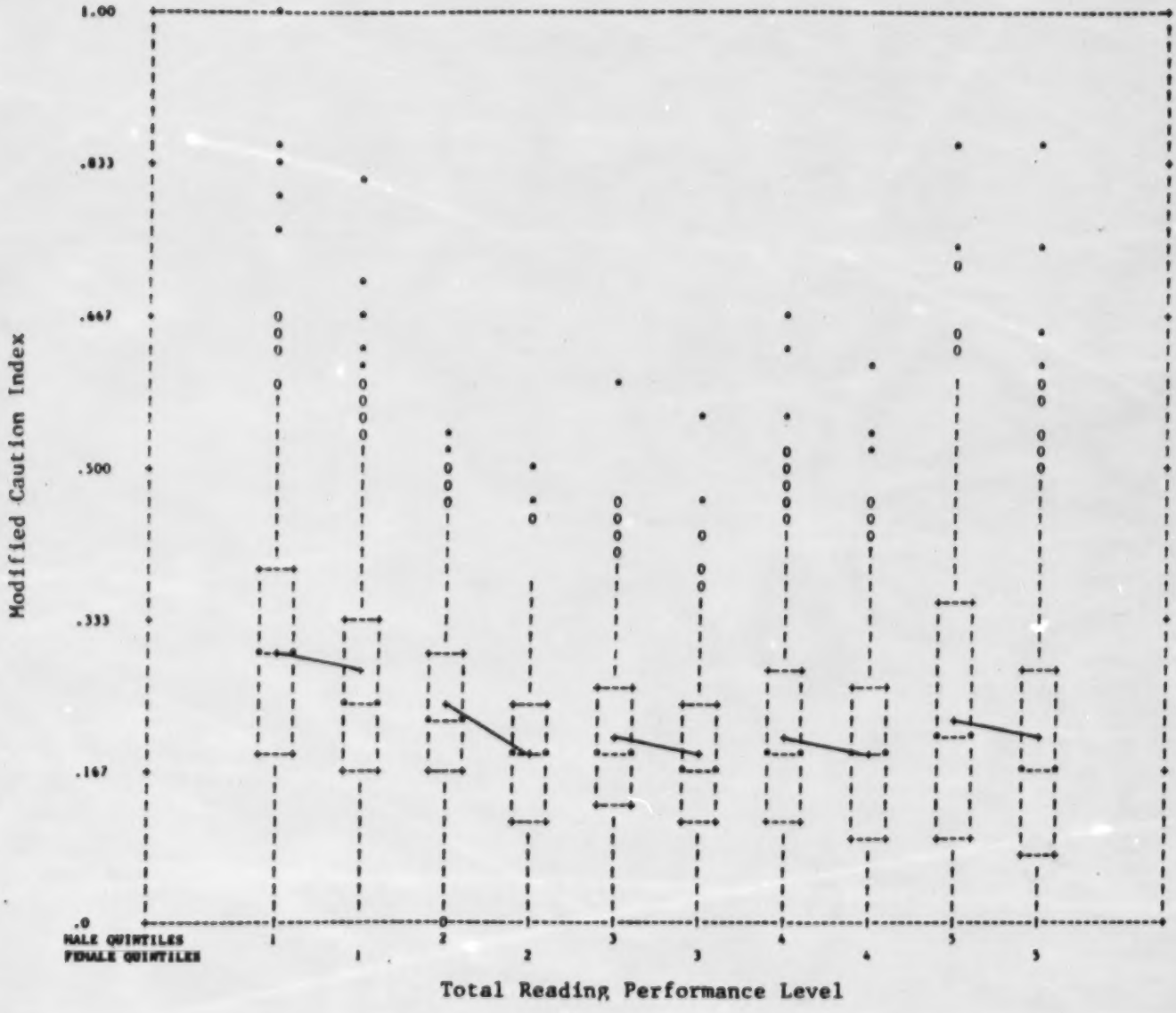

Figure 2. Box and whisker plot of modified caution indices for 9-year-old males and females with the total reading score interval split into quintiles.

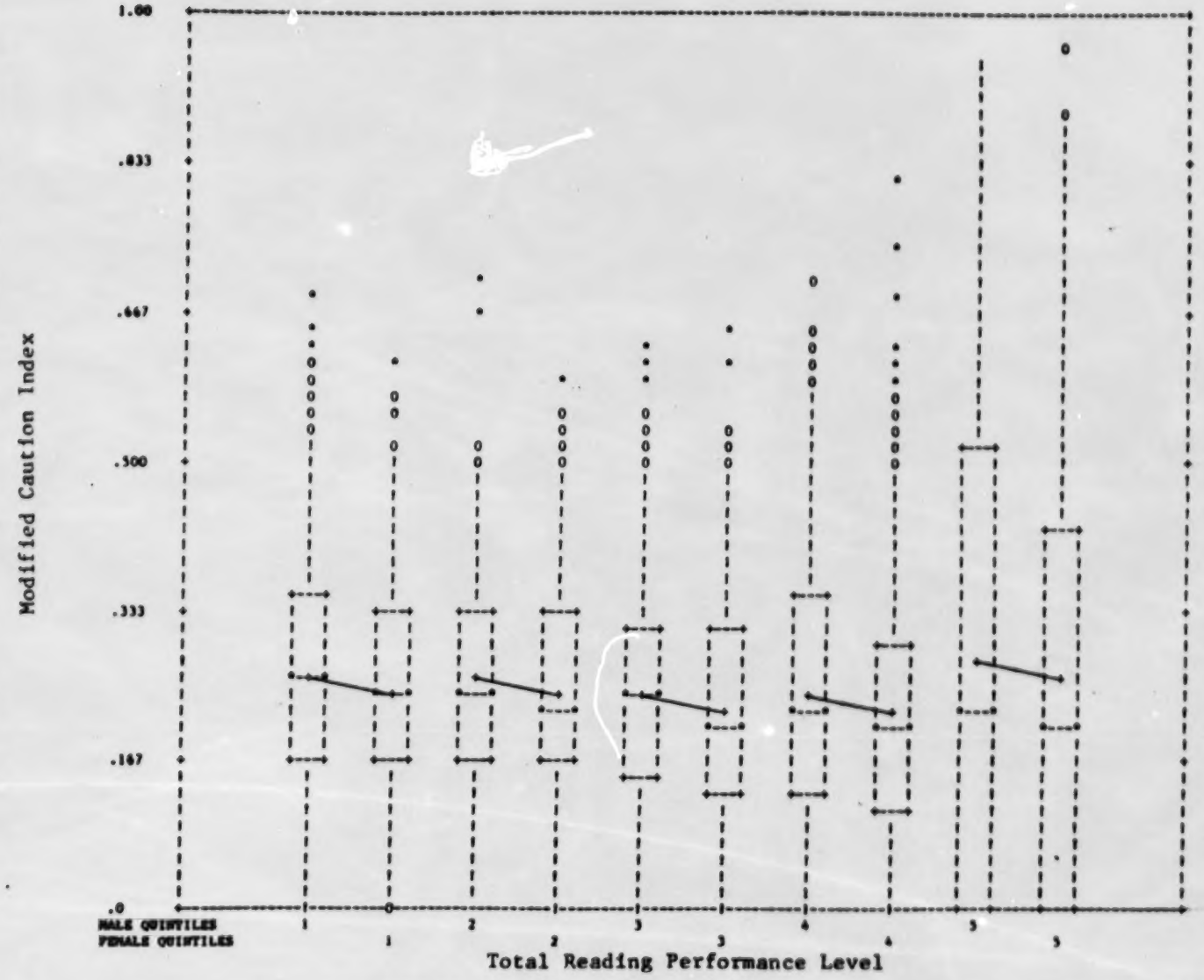

Figure 3. Box and whisker plot of modified caution indices for 13-year-old males and females with the total reading score split into quintiles. (NAEP Booklet 5, Study skills and reading comprehension items, 1979)

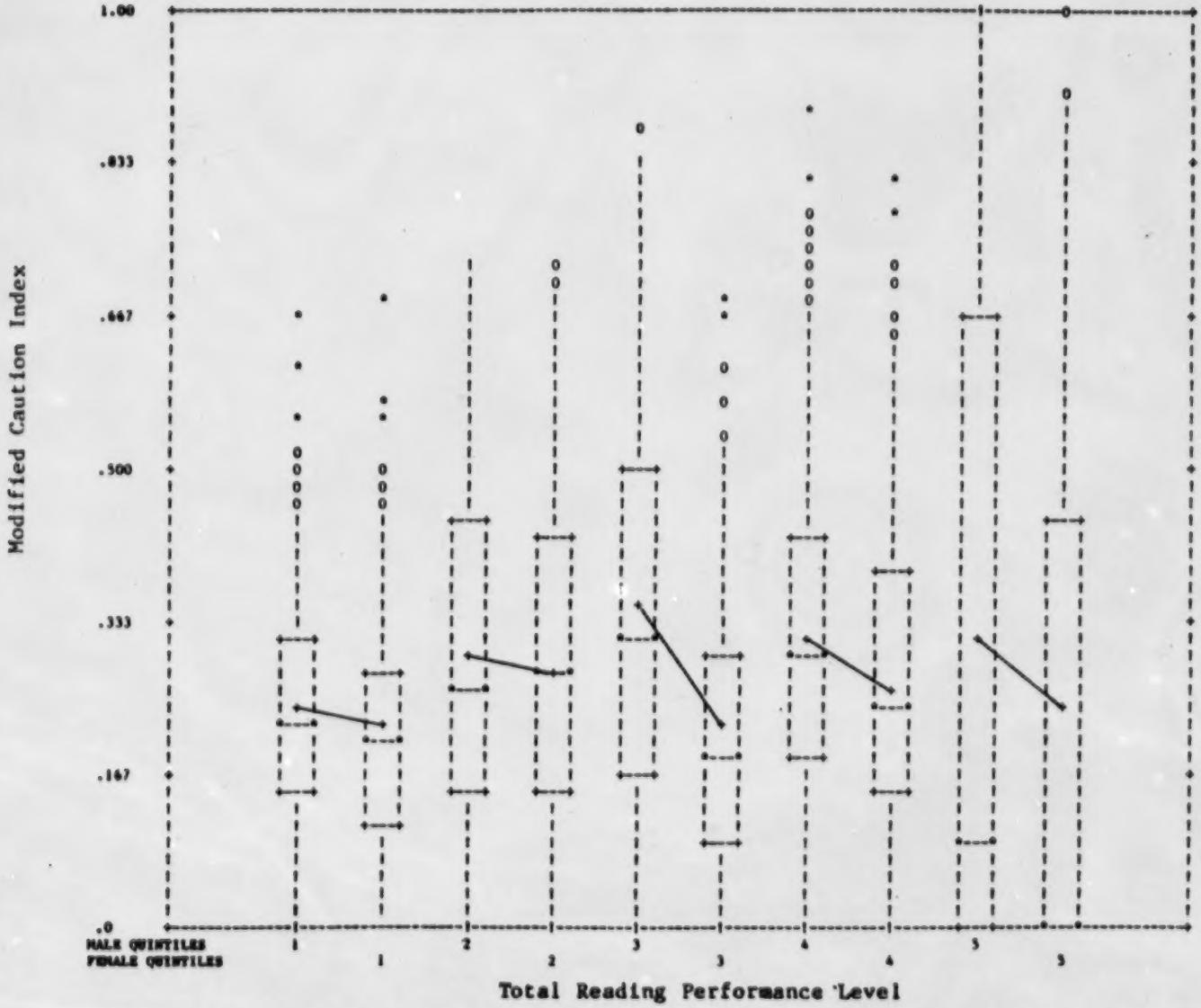

Figure 4. Box and whisker plot of modified caution indices for 17-year-old males and females with the total reading score split into quintiles. (NAEP Booklet 5, Study skills and reading comprehension items, 1979)